Structure, Vol. 13, 347–354, March, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.str.2005.01.012

Visual Methods from Atoms to Cells

Minireview

David S. Goodsell[1,*]
Department of Molecular Biology
The Scripps Research Institute
10550 N. Torrey Pines Road
La Jolla, California 92037

Illustrations of molecular models are widely used for the study and dissemination of molecular structure and function. Several metaphors are commonly used to create these illustrations, and each captures a relevant aspect of the molecule and omits other aspects. Effective tools are available for rendering atomic structures by using several standard representations, and the research community is highly sophisticated in their use. Molecular properties, such as electrostatics, and large complex molecular and cellular systems currently pose challenges for representation.

Browse through any issue of *Structure*, and you will find dozens of pictures of molecules, or more specifically, pictures of models of molecules. In order to create these illustrations, a metaphor must be employed, since we are generating synthetic images of objects that are far smaller than the wavelength of light. The design of these metaphors is the basic challenge of molecular visualization. The most effective metaphors, such as the use of lines to represent covalent bonds or ribbons to represent protein chains, are those that capture relevant aspects of molecular structure and function and make them comprehensible in visual form (Olson and Goodsell, 1992a, 1992b; Richardson, 1992; Hall, 1995).

All metaphors come with the disadvantage of simplification. A molecular model captures only a subset of the properties of the actual molecule. For instance, ribbon diagrams capture the folding of protein chains, but they underestimate the bulk of proteins, whereas space-filling representations show the shape and form of macromolecules, but they often hide all information on bonding and topology. The researcher must carefully choose the representation that captures the property of interest.

Looking through the illustrations in *Structure*, you will find that a few types of illustrations dominate: bond diagrams, ribbon diagrams, space-filling diagrams, and combinations and variations of these. The conventions used in these pictures have become familiar through decades of use. Unless we are told otherwise, lines are automatically understood as atomic bonds, anything colored red is understood to be oxygen, and so on. This article explores the roots of these widely used metaphors, surveys their current use, and presents a few systems in which effective metaphors are still being sought.

*Correspondence: goodsell@scripps.edu
[1]Lab address: http://www.scripps.edu/mb/goodsell

Structure

Most molecular illustrations are illustrations of molecular structure. Molecular structure is naturally amenable to visual metaphors. Molecules are physical objects, with defined sizes and shapes. The vagaries of quantum indeterminacy only become important at submolecular levels (with a few amazing exceptions, such as resonance energy transfer); thus, in many cases, we can treat molecular structures just as we would treat the structure of a house or a chair, by using familiar methods of rendering images of solid objects lit by discrete light sources. Three metaphors—lines, spheres, and ribbons (Figure 1)—have shown lasting success because they each capture an important structural property of the molecule, and they are all easily rendered in a visual form.

The covalent structure of a molecule is effectively displayed through the use of a bond diagram. All of the electrons are discarded, and an artificial bond is created to represent each pair of bonded atoms. This metaphor is highly effective, particularly for organic compounds, because the natural rules of covalent bonding are well defined, with a consistent range of bond lengths, angles, and geometries. The conventions of bond diagrams were codified by early chemists, yielding two common forms: two-dimensional diagrams of organic structures and wood or plastic ball-and-stick molecules used in every organic chemistry classroom. For biomolecules, two representations are commonly used in computer graphics: a single line for each bond or a more sophisticated ball-and-stick model composed of cylinders and spheres. Additional information may be layered onto these representations by coloring the bonds, or by varying the size or texture of cylinders and spheres.

The properties of the electrons are captured in space-filling representations. This representation places a sphere, with a radius that defines the contact distance between atoms, at each atom center. Space-filling representations were developed by Linus Pauling and were made popular through the availability of plastic CPK models (Koltun, 1965). Solvent-accessible surfaces, described in more detail below, are popular variants on the basic approach.

The third widely used representation captures the topology of biomolecules. Both proteins and nucleic acids are composed of a linear chain that folds into a complex three-dimensional structure. Schematics were proposed almost immediately after the first biomolecular structures were solved: the familiar DNA ladder was used in the seminal paper by Watson and Crick (Watson and Crick, 1953), and ribbon diagrams were used by Richard Dickerson to present the structure of myoglobin (Dickerson, 1964). These representations became widely popular after the publication of a survey of protein folding by Jane Richardson (Richardson, 1981), in which she codified a consistent representation for protein secondary structure.

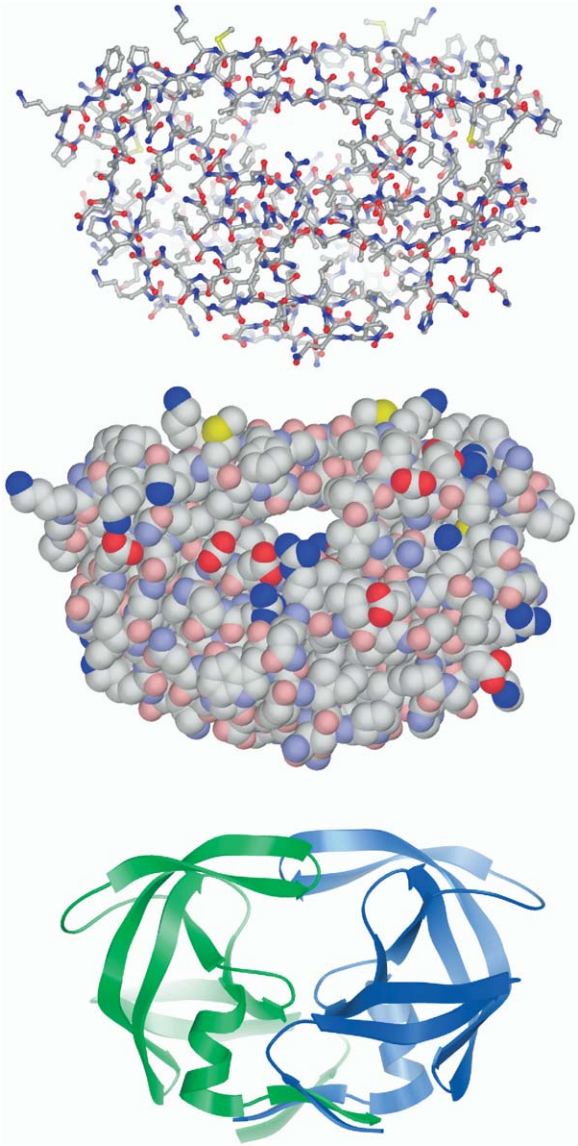

Figure 1. Three Basic Visual Metaphors Are Widely Used to Display Molecular Structure

A bond diagram, shown at the top, uses lines or balls-and-sticks to represent covalent bonds. Traditional colors are used to show the different atom types. A space-filling diagram, shown at the center, uses spheres to show the size and shape of each atom. The coloring scheme is designed to show the chemical nature of each atom while retaining familiar characteristics of the traditional scheme. Charged nitrogen and oxygen are shown in saturated blue and red, respectively, and uncharged nitrogen and oxygen are shown in light blue and pink, respectively. Carbon atoms are shown in white, and sulfur atoms are shown in yellow. A ribbon diagram, shown at the bottom, shows the folding of the two protein chains, one colored blue and the other green. All three diagrams show HIV protease drawn at the same scale. Coordinates were taken from PDB entry 7hvp at the Protein Data Bank (http://www.pdb.org), and the illustration was created by using the Python Molecular Viewer (Sanner, 1999).

Molecular Interaction

Space-filling diagrams are the most general representation for studying the form of molecules. This is one reason that plastic CPK models are so effective: any two space-filling models may be bumped against each other to give a reasonable approximation of dispersion/repulsion forces that define the interaction. The details of this interaction may be refined, however, by defining specialized surfaces that incorporate the features of interaction between molecules.

Solvent-accessible surfaces define the surface of contact between two space-filling representations (Connolly, 1983). A space-filling representation of the protein (or other molecule of interest) is created, and then a probe sphere (most often representing a water molecule) is rolled over the surface. In places where the probe touches, the original space-filling representation is retained. In the cracks and dips between these contact regions, portions of the probe surface are retained, creating a "reentrant surface" that bridges small gaps that are too small for the probe to enter. The surface displays the area that is available for interaction with other atom-sized objects (Figure 2). When combined with color to show the underlying atoms or their properties, solvent-accessible surfaces are an effective tool for presenting molecular interactions.

A volume-based method takes a different approach to the interaction. Instead of rolling the probe over the surface, the probe is scanned through the entire three-dimensional space in and around the molecule. At each location, the interaction energy of the probe is saved (Goodford, 1985; Goodsell and Olson, 1990). The result is a map of favorable regions and unfavorable regions that may be displayed by using the volumetric metaphors described in the next section (Figure 2). Notice the similarity between these two approaches: both display the spatial extent of the interaction between a macromolecule and a probe. The difference is in the perspective: the volume-based method shows the interaction from the perspective of the probe, whereas solvent-accessible surfaces display the interaction from the perspective of the macromolecule.

Volumetric Properties

The volumetric properties of molecules—the properties that extend through space within and around molecules—do not have such direct macroscopic analogs, and we are still designing metaphors to capture and present these properties. Electron density distributions, from X-ray crystallography or electron microscopy, are the most common volumetric data sets used by structural molecular biologists. These are relatively simple to approach, since they typically have a single scalar value at each point in space. However, the properties of fields, such as electrostatics or magnetism, pose greater challenges, since these fields vary in magnitude and direction over a volume of space. The choice of a metaphor can be tricky. Fields are not particularly prevalent in macroscopic life, so we have few familiar methods for presenting them. However, they are critically important at the molecular level. A few simple metaphors have been widely employed for scalar volumetric data sets (Figure 3), but researchers are still searching for effective methods for displaying the complex, spatially varying properties of fields (Figure 4).

The concept of an isovalue surface is widely used for representation of volumetric properties, in particular, the electron distributions that result from X-ray crystal-

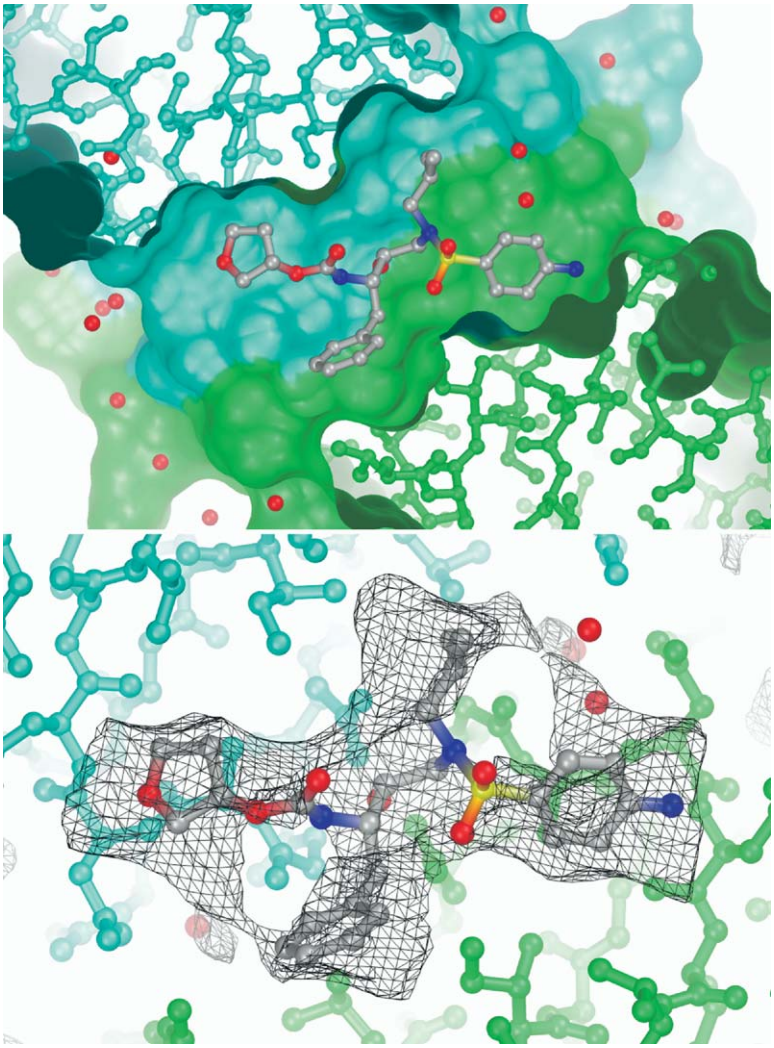

Figure 2. Solvent-Accessible Surfaces and Interaction Potentials Show the Interaction of an Inhibitor with HIV-1 Protease

In both images, the protein atoms are displayed with a ball-and-stick representation, with one chain in blue and the other chain in green. The inhibitor is shown in ball-and-stick, colored with traditional atomic coloration. Water molecules are shown with small red spheres. In the top image, a solvent-accessible surface of the protein is shown and is clipped to show the hourglass shape of the active site tunnel. The lower image shows an interaction potential (Goodsell and Olson, 1990) of the same active site. The cage of lines encloses the region where inhibitor atoms have highly favorable interaction energies with the protein, and, as expected, the inhibitor falls within this region. Coordinates are from entry 1hpv at the Protein Data Bank, and the illustrations were created with the Python Molecular Viewer.

lography and electron tomography. A surface is drawn to enclose all regions with values higher than a given threshold. These representations are effective and widely used because surfaces are easy to visualize and the concept of an enclosing boundary is easy to comprehend.

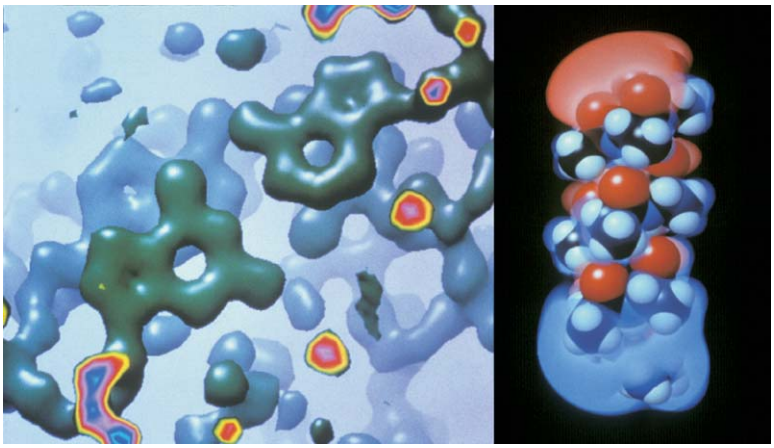

Figure 3. Volumetric Properties Are Commonly Displayed with Two Methods

At the left, a crystallographic electron density distribution is displayed by using isocontours. The green surfaces surround regions of high electron density, corresponding to atoms in this DNA oligonucleotide (notice the unusual guanine-adenine base pair at center). The data space is clipped in this image, and the cross-section is colored by the local value of the electron density. Low values are transparent, but the regions of high electron density are colored yellow to red to magenta at increasing values. The right image uses a cloudy voxel representation to display the electrostatic potential around an α helix. Areas with strong negative potential are shown in red, and areas with strong positive potential are shown in blue. Notice how alignment of the peptide groups forms a large dipole across the entire helix.

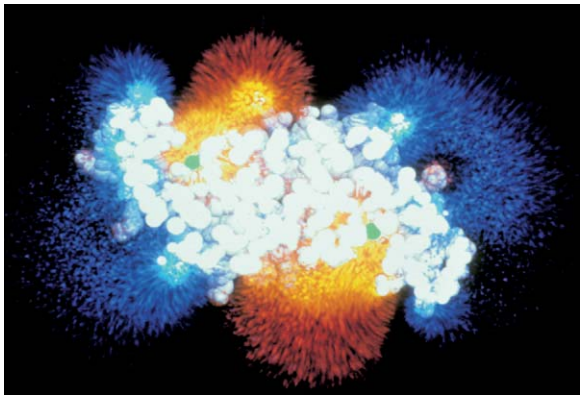

Figure 4. Volumetric Properties Pose Challenges for Design of Visual Metaphors

This image is an experiment in using textures to display the direction and magnitude of the electrostatic field around superoxide dismutase. A cross-section through the protein is shown, cutting through the two active sites. Regions of positive electrostatic potential are shown in red and yellow, and regions of negative electrostatic potential are shown in blue. The texture shows the local direction of the electrostatic field, which is thought to steer negatively charged superoxide molecules from the blue regions here into the red regions, ultimately steering the superoxide into the copper and zinc ions shown in green. Coordinates were taken from entry 2sod from the Protein Data Bank.

Figure 5. Solid Models of Molecules Are Being Created by Using Rapid Prototyping Technologies

This model of chymotrypsin was created automatically with the Z-corporation technology, which lays down thin layers of gypsum powder and then applies colored glue with an ink-jet printer. The result is a solid model created according to the experimental atomic coordinates (taken from entry 4cha at the Protein Data Bank).

The second metaphor is a volumetric metaphor, in which each location in space is assigned optical characteristics, such as color and opacity, based on the local value of the property (Goodsell et al., 1989). Two methods of creating images from this concept are widely used. The first creates a cross-section of the space. A plane is placed in a strategic spot and is colored by the local value of the continuous property. This is an effective metaphor, since we are familiar with cutting objects in half and looking at the distribution of textures and colors on the exposed surface.

The second approach uses a cloudy representation, using small, semi-opaque voxels at each point. Early experiments, such as that shown in Figure 3, show the possibilities of the method. The method has been limited, however, by the magnitude of the computation and difficulty in assigning effective transfer functions (which determine the color and opacity of different data values). Over the past decade, dedicated hardware combined with research into automatic assignment of transfer functions is promising to make this method a practical tool.

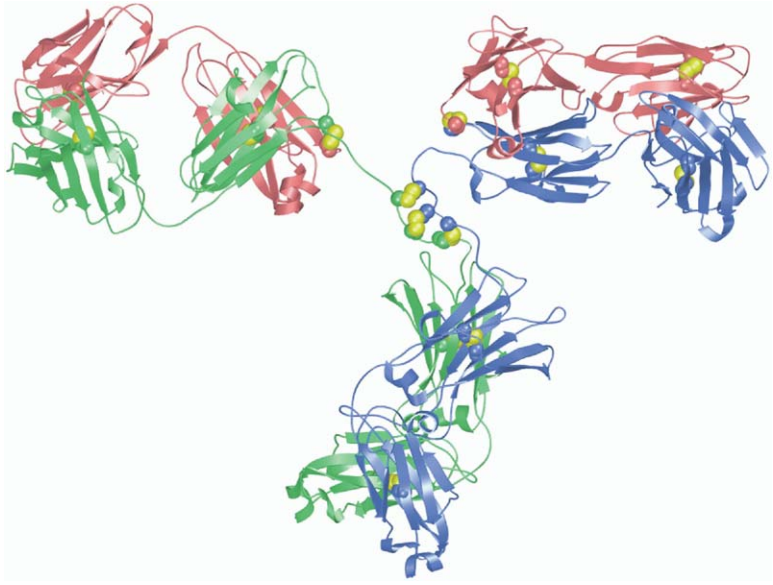

Figure 6. Ribbon Diagrams Are the Most Common Type of Illustration Used in Structural Research Articles

In this illustration, the four chains of an antibody are shown with ribbons, and the many disulfide linkages within and between chains are shown with spheres. The two light chains are in red, and the two heavy chains are in blue and green. Coordinates were taken from entry 1igt at the Protein Data Bank, and the illustration was created with the Python Molecular Viewer.

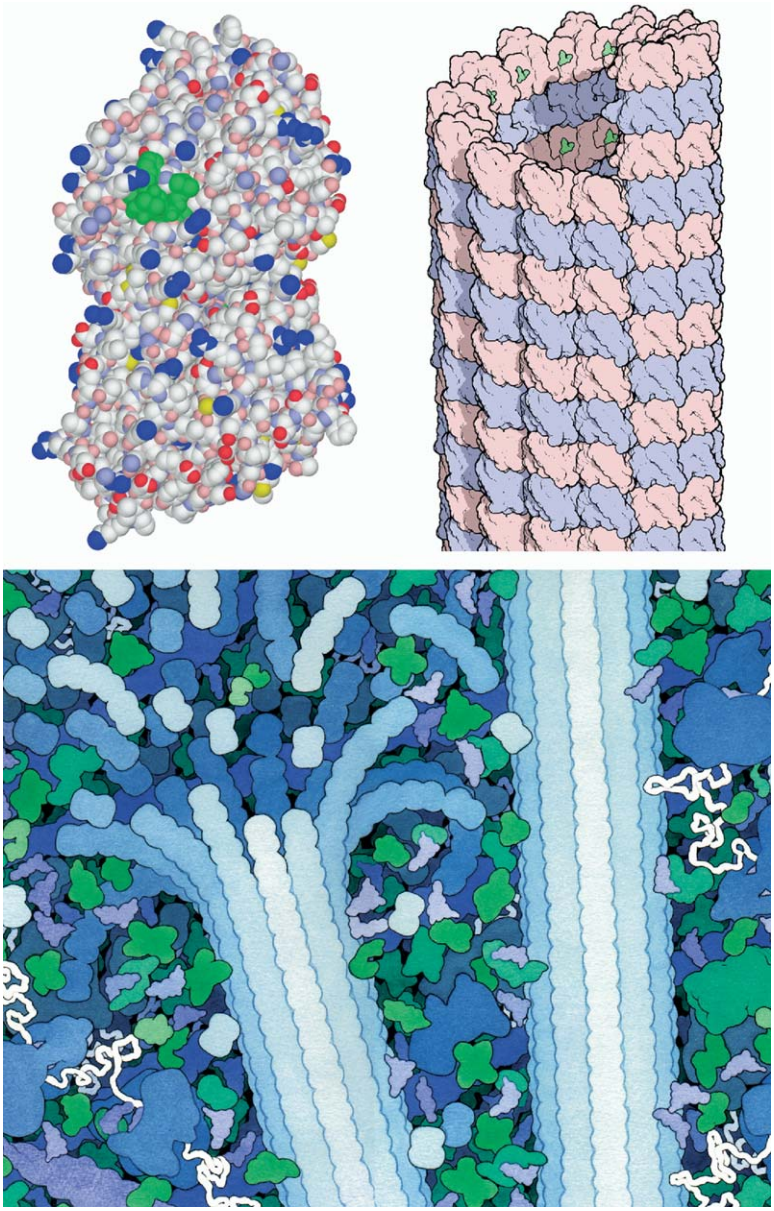

Figure 7. A Consistent Space-Filling Style Can Help with the Transition between Scale Levels

At the top left, a space-filling diagram shows one dimer of tubulin, with the drug taxol shown in green. At the top right, the stacking of tubulin in a microtubule is rendered with outlines and flat colors, but still retains the space-filling shape of each dimer. At the bottom, a hand-drawn illustration simplifies the outlines and colors, showing microtubules in their cellular context. Coordinates were taken from entry 1tub at the Protein Data Bank.

Rendering

Once a metaphor is chosen, all of the tricks of the artist may be used to make this visual metaphor comprehensible. Visual representation is, in many ways, a solved problem, at least for scenes composed of surfaces. Decades of work by illustrators and researchers have explored many possible modes of casting a given model into a visual representation.

Effective computer graphics methods are available for any representation composed of surfaces. These allow rendering of images that simulate all manner of colors and textures by using a variety of lighting models. The most popular modes, often provided as the default rendering choices in widely available software, render molecular models as shiny surfaces with saturated colors, giving the impression of a photograph of a plastic model. These are highly effective because they take advantage of our natural ability to recognize highlights, shadows, and occluded surfaces as clues to comprehend the three-dimensional form of an object.

Illustrative methods are also available to simplify renderings. The earliest software for molecular rendering, such as ORTEP and PLUTO, used pen plotters to create images, producing illustrative images composed of outlines and lines of intersection. Jane Richardson's beautiful hand-drawn ribbon diagrams (Richardson, 1985) lead to the development of Molscript (Kraulis, 1991), which automatically creates cartoon images of ribbon diagrams. Image-processing techniques (Namba et al., 1989; Goodsell and Olson, 1992) may also be used to create illustrative outlines in molecular structures. All of these methods have the advantage of simplifying the object being displayed by using the artificial outlines and shading lines to enhance the description

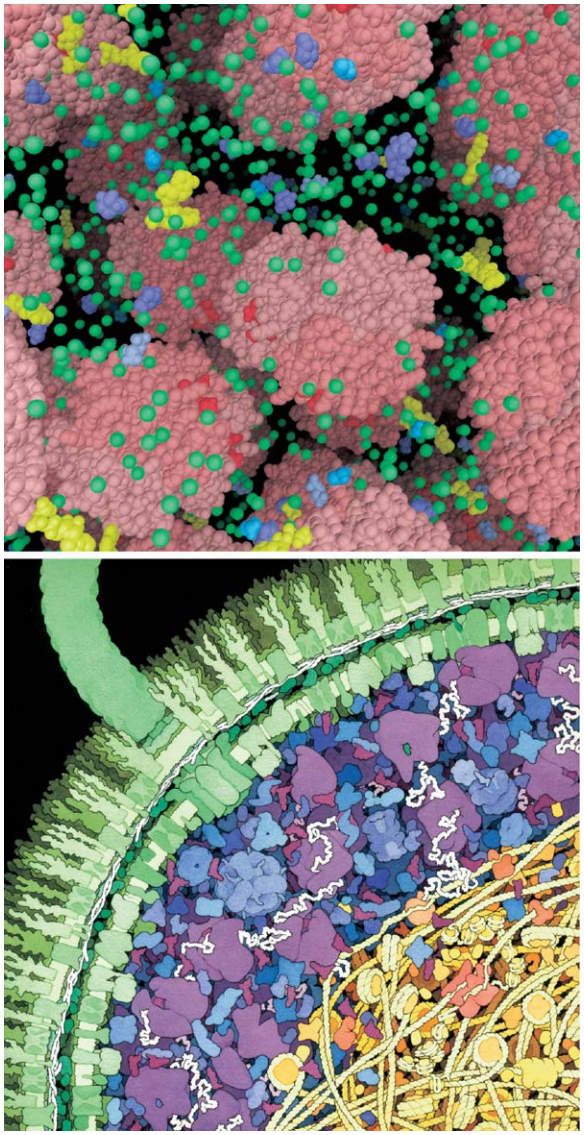

Figure 8. Large Systems May Be Displayed with an Immersive Approach or as a Cross-Section

At the top, the interior of a red blood cell is shown in an immersive image, with hemoglobin in red, small molecules in blue and yellow, and ions in green. The camera is placed within the scene so the viewer is surrounded by molecules. At the bottom, a cross-section through an *Escherichia coli* cell is shown, including all macromolecules. The double-layered cell wall is shown in green, with a flagellum extending upward to the left. The cytoplasmic space is shown in blue, with ribosomes in purple. The nucleoid area is shown at the lower right, with DNA strands in yellow.

of the form. They are also amenable to printing as line art in black and white.

Looking to the future, we are now seeing a return to physical models of molecules. Rapid prototyping techniques, which build solid models layer by layer, are now becoming available to create solid models from surface-based representations of molecules (see the article by Olson in this issue), as shown in Figure 5. Now, instead of tricking the eye into seeing the three-dimensional form of a model of a molecule, we can interact directly with the molecular model.

Molecular Graphics in the Laboratory

Today, fast computer graphics methods are available for creating all of the basic representations and their variants. Dozens of turn-key programs are available for converting a set of atomic coordinates into an image (see, for instance, the list at the Protein Data Bank: http://www.pdb.org/links.html), and computer graphics are widely used to disseminate scientific results. To get a representative sample of current use, I surveyed the figures presented in 103 articles published in *Structure* from July 2004 to December 2004, and I found that the research community is highly sophisticated when it comes to using these tools.

When representing entire protein structures, ribbon diagrams are by far the most commonly used representations (Figure 6) for subjects ranging all the way from small hormones to virus capsids. Of the 103 papers, 67 include ribbon diagrams of monomeric and dimeric proteins, and 13 papers include ribbon diagrams of larger oligomers. A simpler backbone representation, using a smooth tube or bonds that connect α carbon positions, was also widely used in 24 papers that showed monomeric or dimeric proteins and 7 papers presenting larger structures. In about half of these ribbon and backbone pictures, selected ligands or side chains were also included; ball-and-stick representations were used in 31 cases, and space-filling representations were used in 15 cases. In cases in which several overlapped structures were shown, the figures where evenly divided. Ribbon diagrams were used in 31 cases, most often when only 2 or 3 structures were overlapped, and 30 illustrations used simpler backbone traces, typically in illustrations that had 10 or more structures overlapped in a single image.

Traditional space-filling representations of entire proteins were used in only five papers, usually in cases in which sequence conservation was discussed. However, smooth surfaces colored by electrostatics or other properties were presented in 24 papers, presumably due to wide availability of the program GRASP (Nicholls et al., 1991).

Three-quarters of these papers also included one or more close-up illustrations of functionally important portions of proteins. These illustrations use a wide variety of techniques to show the atomic details, often combining ribbons, tubes, ball-and-stick, space-filling, and surface representations in a single image. The most common type of close-up image includes the protein backbone as a tube representation and displays side chains and ligands in ball-and-stick representation. This is an effective combination of methods for focusing attention on the interaction while still allowing the viewer to recognize the underlying chemical structure of relevant parts of the molecules.

When we move to larger structures, such as the ribosome or viral capsids, we find researchers using the same techniques that are used with smaller structures. Structural reports will typically include one figure that shows the whole complex, with each of the protein and nucleic acid chains shown in a ribbon or tube representation. Color and depth cuing are used to improve the comprehensibility of these complicated illustrations,

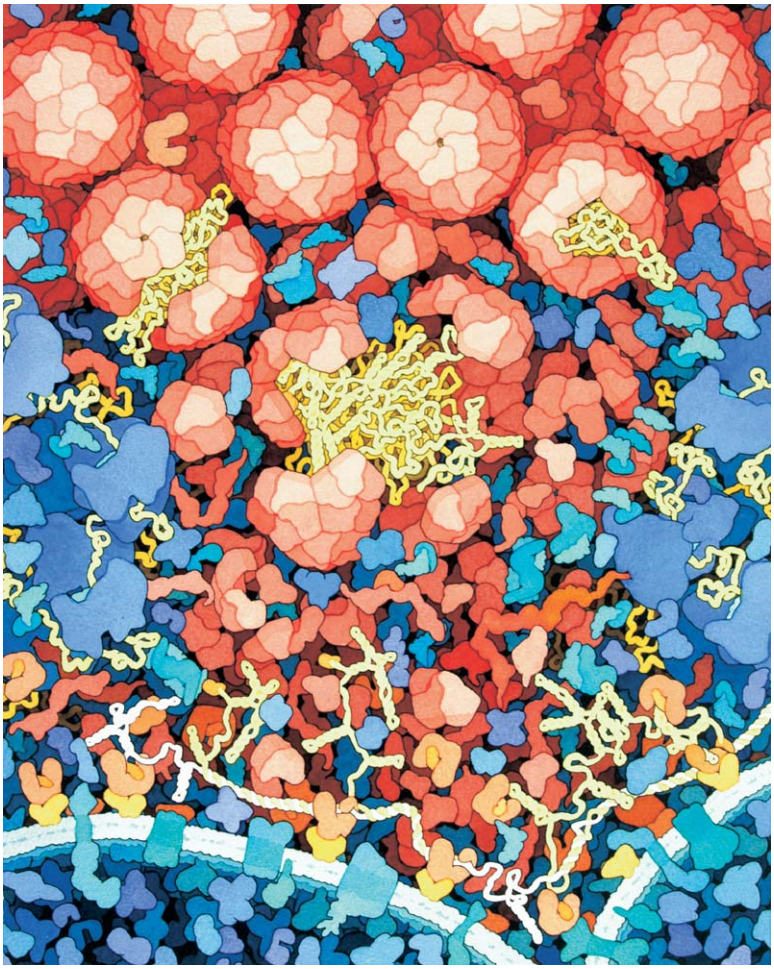

Figure 9. Poliovirus Self-Assembly Is Shown in a Cross-Section through an Infected Cell

The cross-section metaphor allows each step of this process to be shown in a single illustration. Synthesis of RNA by viral polymerases is shown at the bottom, with the minus strand in white and new plus strands in yellow. At the center, ribosomes (in blue) are shown synthesizing viral proteins (in red), which then assemble with the RNA to form viruses at the top.

and a hand-drawn schematic drawing is often included to help present the geometry of the various subunits. Then, a series of illustrations typically show the features of interest, highlighting parts of the complex in more detail or dissecting out individual subunits for separate display.

From Molecules to Cells

Molecular biology is gradually merging with cell biology as data become increasingly available over the entire scale range of nanometers to millimeters. In several well-characterized systems, such as red blood cells and *Escherichia coli* cells, it is possible to construct convincing models of portions of cells, showing the concentration and distribution of macromolecules (Goodsell, 1991, 1992a, 1992b). These complex systems pose new challenges for representation. In my own work, I have used two concepts to make these complex systems more readable.

The first concept is the use of a consistent hierarchical style across scale (Goodsell, 2000), as shown in Figure 7. The underlying metaphor simulates what we might "see" if the object were enlarged to visible size. The space-filling representation is used, since it represents the shape and size of each molecule. The use of a space-filling approach allows the combination of close-up pictures, which show the atomic details of each molecule, with larger fields of molecules, where atomic detail would add too much complexity. By progressively smoothing the representation as larger and larger fields are shown, the image maintains an appropriate level of comprehensibility at each scale level. The similarity in shape between the different levels allows the viewer to move from one image to the next and identify individual molecules.

The second concept is the use of a cross-section metaphor instead of an immersive metaphor in depictions of cell environments, as shown in Figure 8. Immersive images place the viewer within the field of objects. They are dynamic, but they are subject to the distortions of the perspective transformation and occlusion by the nearest objects. The cross-sectional metaphor allows the display of large fields of molecules and entire molecular processes (Figure 9) in a way that would be impossible with the immersive approach. In addition, the cross-sectional metaphor is familiar in this context, since electron and light micrographs, which are just a step or two lower in resolution than the illustrations, are typically images of cellular cross-sections.

Perspective

Molecular graphics is a mature discipline, with a wide variety of methods in common use by researchers, educators, and students. Many of the challenges currently

under study are challenges of magnitude: how to search through and display databases of hundreds of structures, how to deal with structures composed of millions of atoms, and how to model and analyze systems of thousands of individual molecules. These large applications will require advances at all levels. They will require experimentation into the conventions used for representation to allow comprehensible display as the systems get more and more complex. They will require new definitions in community standards to ensure that our databases can encompass the growing magnitude of these systems. And finally, they will depend on continued advance of computational capability to allow timely interaction with the data.

Acknowledgments

The author kindly acknowledges support from the Protein Data Bank. This is publication 17015-MB from the Scripps Research Institute.

Selected Reading

Connolly, M.L. (1983). Science *221*, 709–713.

Dickerson, R.E. (1964). In The Proteins, Volume 2, H. Neurath, ed. (New York: Academic Press), pp. 634.

Goodford, P.J. (1985). J. Med. Chem. *28*, 849–857.

Goodsell, D.S. (1991). Trends Biochem. Sci. *16*, 203–206.

Goodsell, D.S. (1992a). Am. Sci. *80*, 457–465.

Goodsell, D.S. (1992b). The Machinery of Life (New York: Springer-Verlag).

Goodsell, D.S. (2000). J. Biocommun. *27*, 12–18.

Goodsell, D.S., and Olson, A.J. (1990). Proteins *8*, 195–202.

Goodsell, D.S., and Olson, A.J. (1992). J. Mol. Graph. *10*, 235–240.

Goodsell, D.S., Mian, I.S., and Olson, A.J. (1989). J. Mol. Graph. *7*, 41–47.

Hall, S.S. (1995). Science *267*, 620–624.

Koltun, W.L. (1965). Biopolymers *3*, 665–679.

Kraulis, P. (1991). J. Appl. Crystallogr. *24*, 946–950.

Namba, K., Caspar, D.L.D., and Stubbs, G. (1989). Biophys. J. *53*, 469–475.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Proteins *11*, 281–296.

Olson, A.J., and Goodsell, D.S. (1992a). Curr. Opin. Struct. Biol. *2*, 193–201.

Olson, A.J., and Goodsell, D.S. (1992b). Sci. Am. *267*, 76–81.

Richardson, J.S. (1981). Adv. Protein Chem. *34*, 167–339.

Richardson, J.S. (1985). Methods Enzymol. *115*, 359–380.

Richardson, J.S. (1992). Biophys. J. *63*, 1186–1209.

Sanner, M.F. (1999). J. Mol. Graphics Mod. *17*, 57–61.

Watson, J.D., and Crick, F.H.C. (1953). Nature *171*, 737–738.